



U.S. Department of the Interior
Bureau of Land Management



Influence of Benchmark Method Accuracy and Precision on Assessments of Aquatic Resource Condition

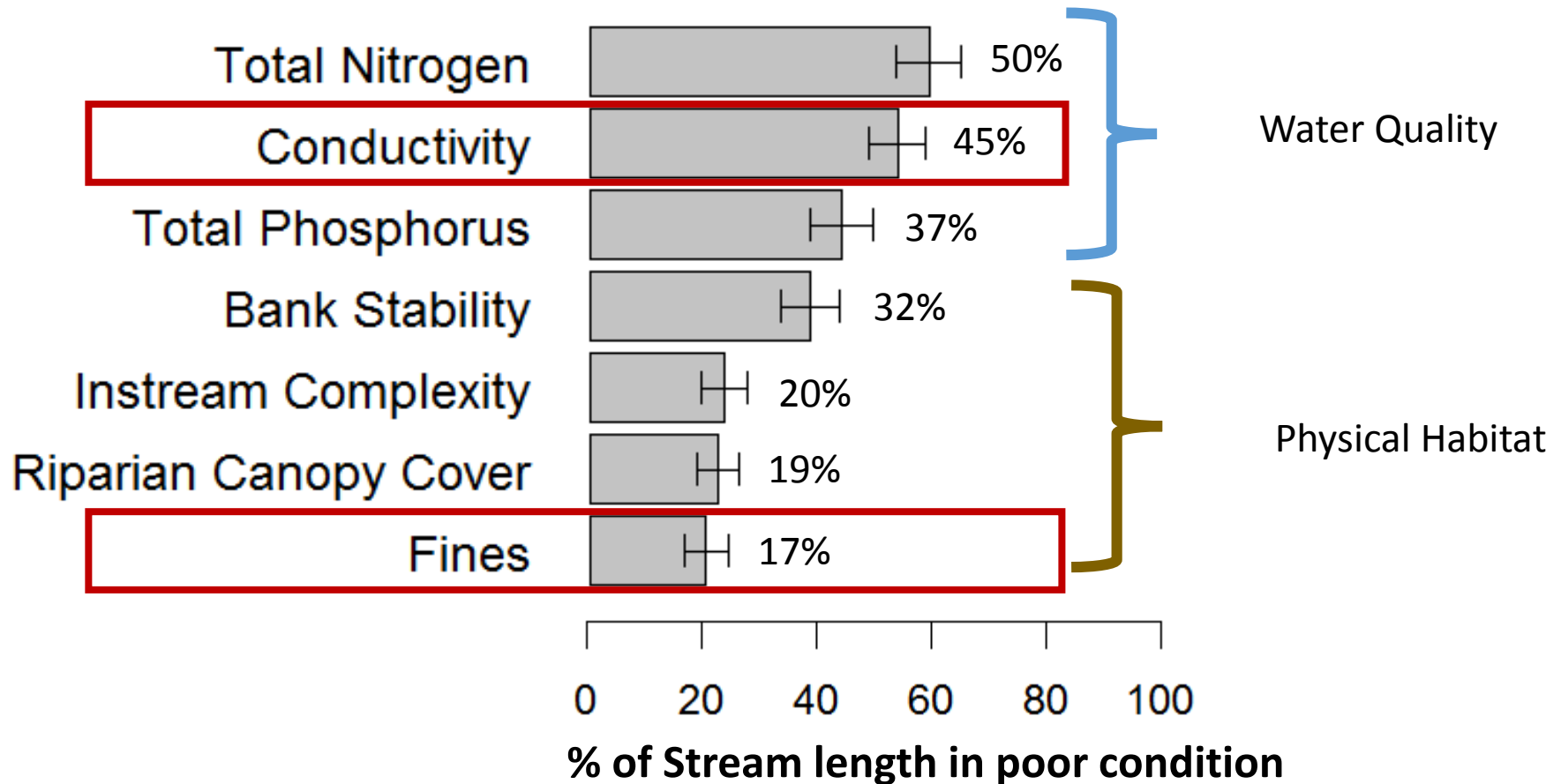


Jennifer Courtwright¹, Nicole Cappuccio²,
Scott W. Miller^{1,2}, and Charles P. Hawkins¹

¹ BLM/USU National Aquatic Monitoring Center (aka The BugLab)

² BLM-National Operations Center, Branch of Assessment and Monitoring

Ecological Assessments



Different benchmark methods confound comparisons among indicators!

Talk Outline

- Compare and contrast common benchmark methods for conductivity and fine sediment
- How do the concepts of accuracy and precision apply to benchmark methods?
- How do we rank stressors given differences in benchmark method accuracy and precision across indicators and agencies?

Talk Outline

- Compare and contrast common benchmark methods for conductivity and fine sediment
- How do the concepts of accuracy and precision apply to benchmark methods?
- How do we rank stressors given differences in benchmark method accuracy and precision across indicators and agencies?

Methods for Setting Benchmarks

- Site-specific empirical models – use networks of reference sites to set site-specific benchmarks across continuous natural gradients
- Percentiles of regional reference – use networks of reference sites to set benchmarks by landscape class
- Biological response – set benchmarks at values that correlate with a change in biological condition
- Best professional judgement – use literature and professional experience to set benchmarks

Example Benchmark Methods Across Agencies

Indicator	Method	Agency
Conductivity	Model (Olson and Hawkins 2012)	BLM
	Best professional judgement	EPA and most states
% Fines	Percentiles of regional reference	BLM
	RBS Model (Kaufmann et al. 2018)	EPA
	Biological response	States

Example Benchmark Methods Across Agencies

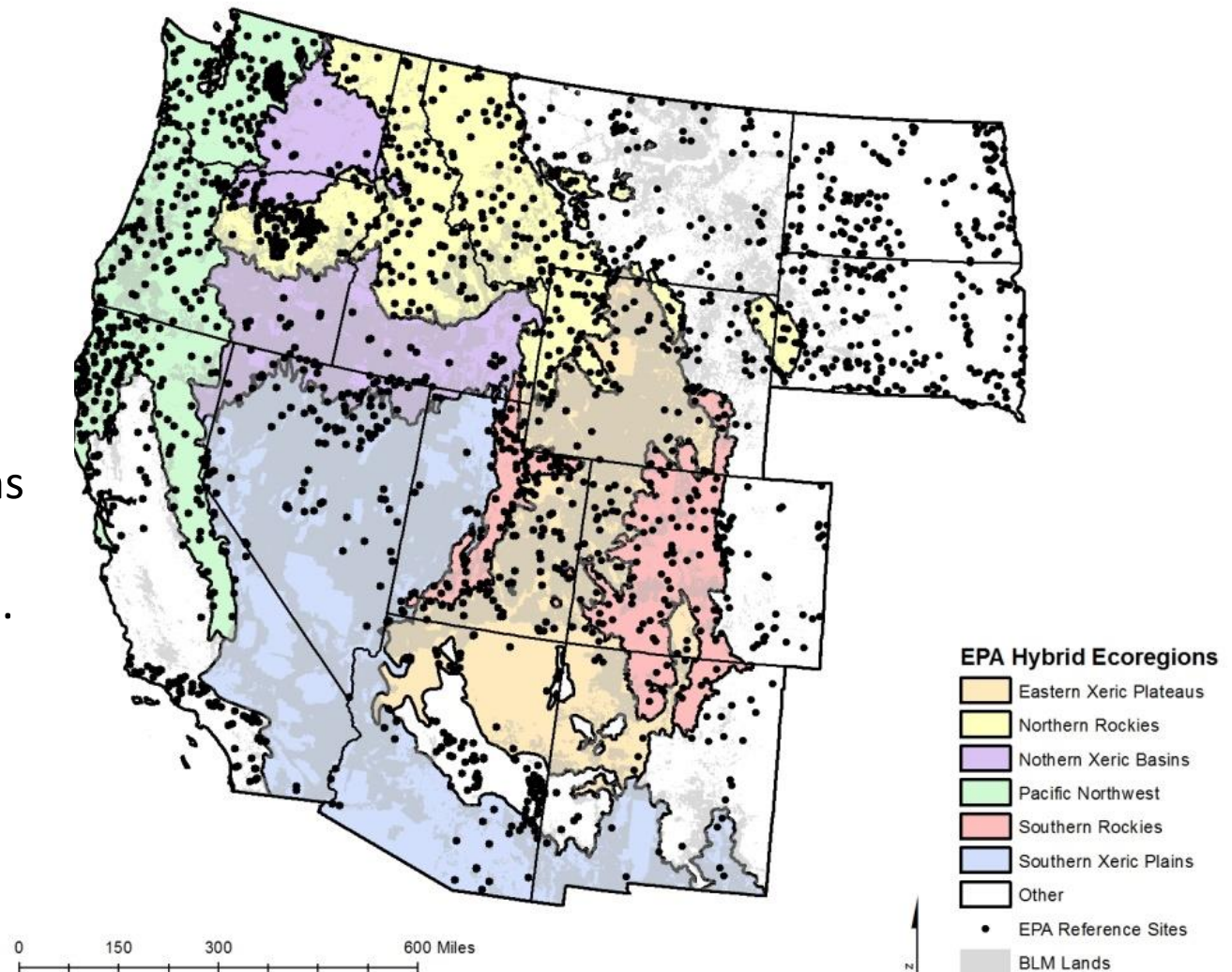
Indicator	Method	Agency
Conductivity	Model (Olson and Hawkins 2012) Best professional judgement	BLM EPA and most states
% Fines	Percentiles of regional reference RBS Model (Kaufmann et al. 2018) Biological response	BLM EPA States

Interagency Networks of Reference Sites

Stoddard et al. 2006

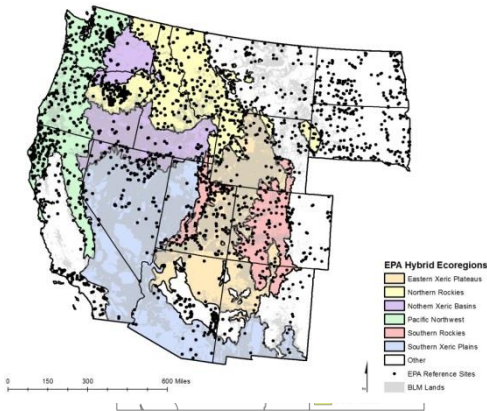
Sites screened for:

- Road density
- Timber harvest
- Grazing
- Dams
- Artificial flow paths
- % Agriculture
- Other land-uses.....

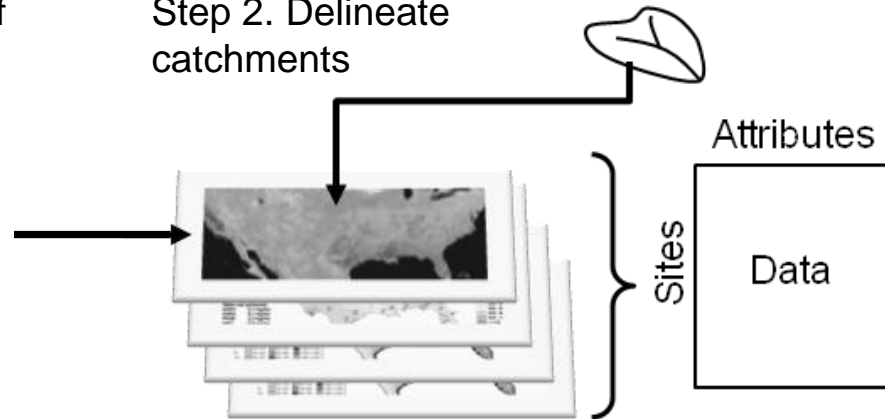


Site Specific Empirical Models: Conductivity

Step 1. Sample network of reference sites

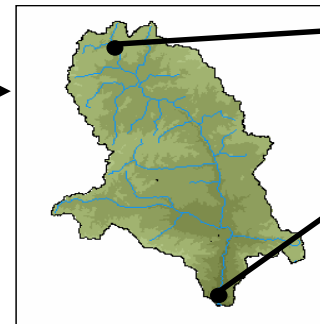
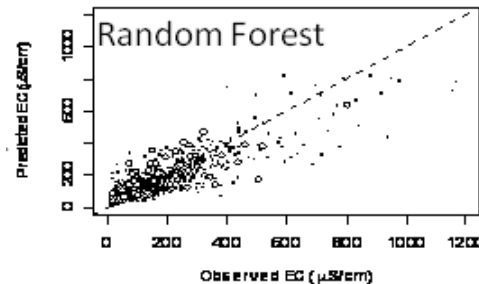


Step 2. Delineate catchments



Step 3. Compute catchment statistics for geology, climate, vegetation, soils, topography, etc.

Step 4. Build reference site models



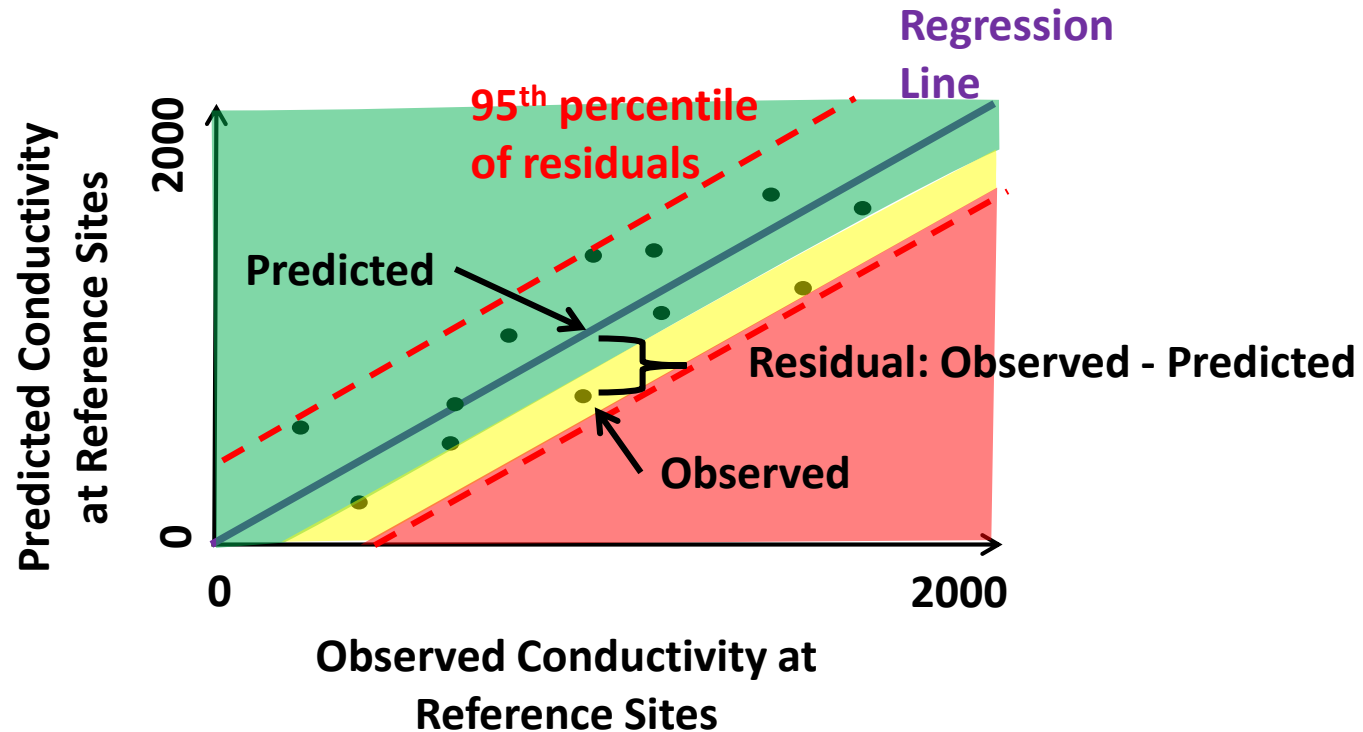
Observed: 120 $\mu\text{S/cm}$

Predicted: 125 $\mu\text{S/cm}$

Observed: 200 $\mu\text{S/cm}$

Predicted: 175 $\mu\text{S/cm}$

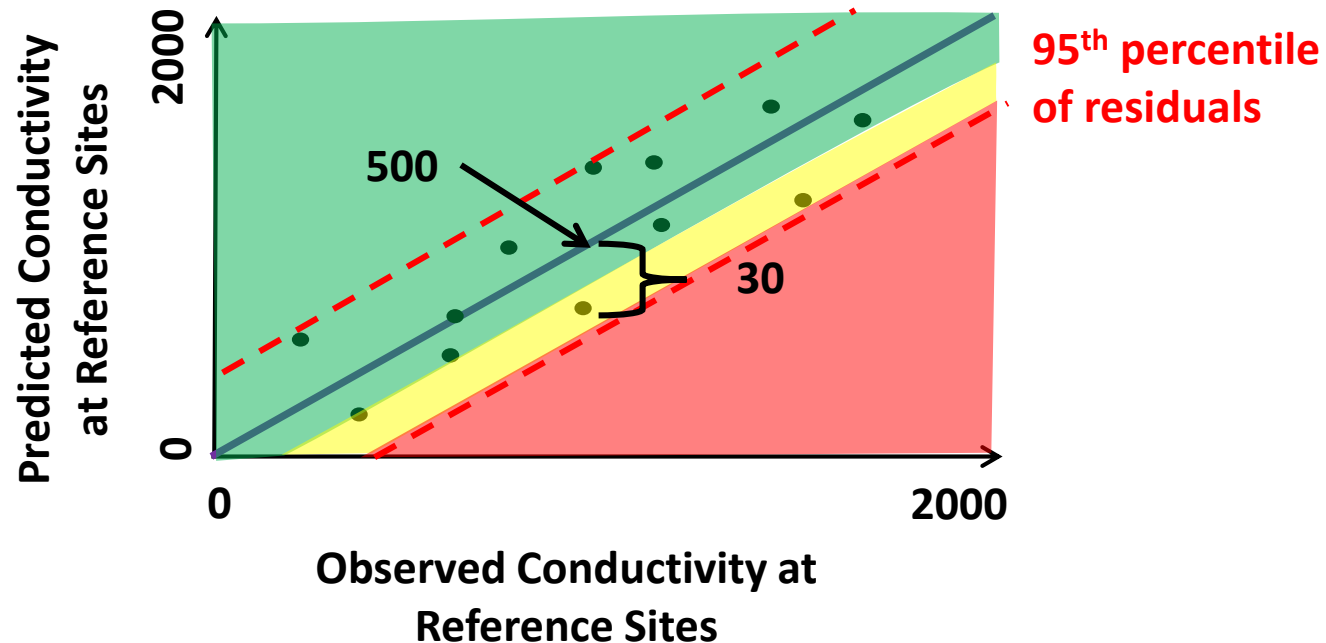
Site Specific Empirical Models: Conductivity



Fair: Predicted value + 75th percentile of residuals

Poor: Predicted value + 95th percentile of residuals

Site Specific Empirical Models: Conductivity

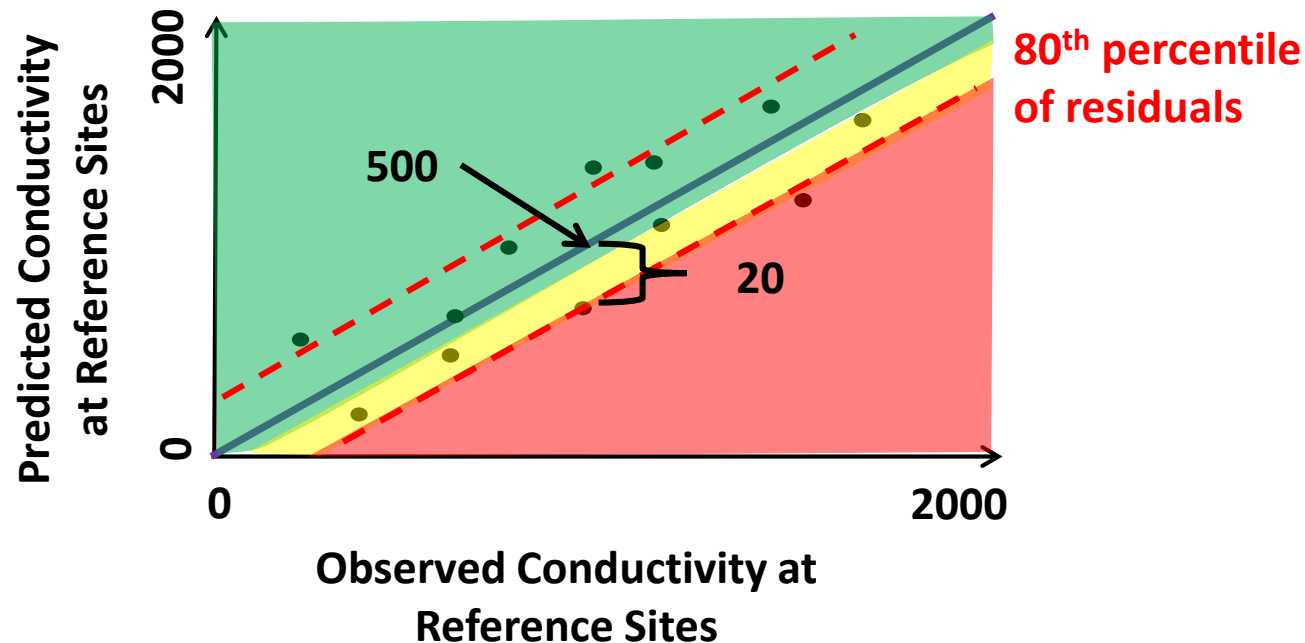


Percentiles chosen based on best professional judgement

The lower the percentile the more protective benchmark

- 95th percentile: 30 Poor: $500 + 30 = 530 \mu\text{S}/\text{cm}$

Site Specific Empirical Models: Conductivity



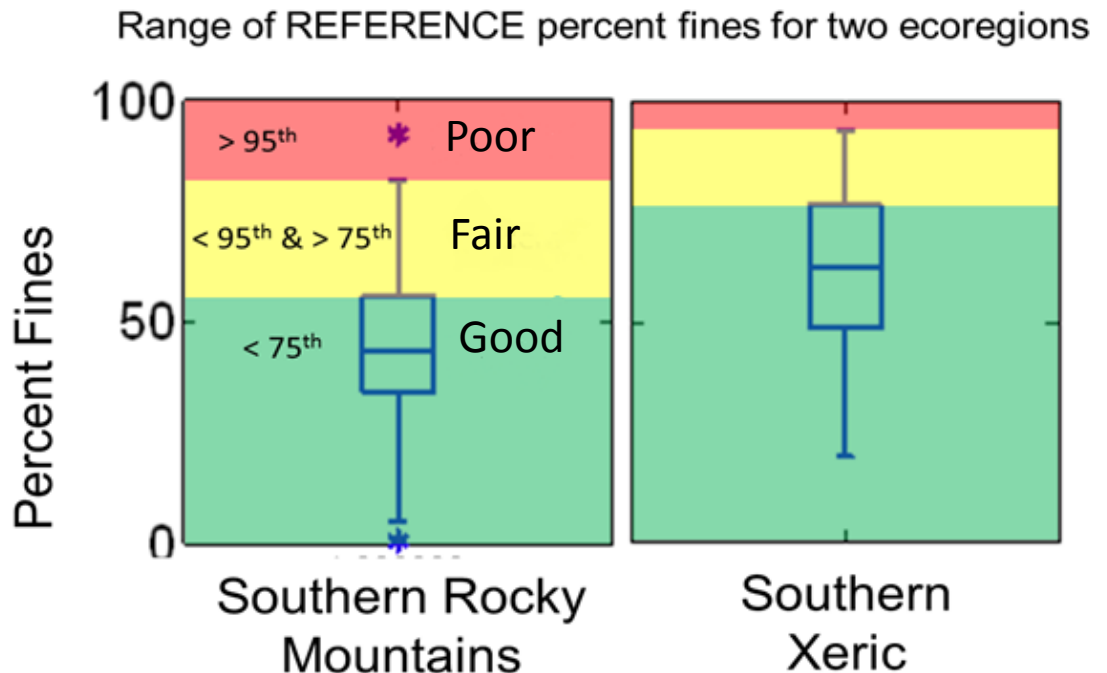
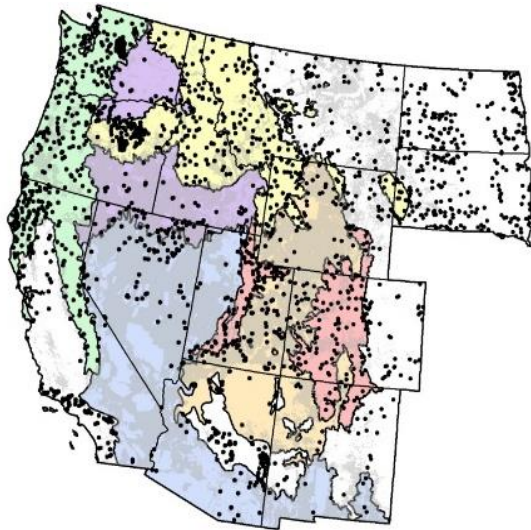
Percentiles chosen based on best professional judgement

The lower the percentile the more protective benchmark

- 95th percentile: 30 Poor: $500 + 30 = 530 \mu\text{S/cm}$
- 80th percentile: 20 Poor: $500 + 20 = 520 \mu\text{S/cm}$

Percentiles of Regional Reference: % Fines

- A reference network is sampled
- Categorical methods are used to partition natural variance
- Percentiles chosen based on best professional judgement
 - The lower the percentile the more protective the benchmark
 - Poor: 95th percentile = 80% fines Southern Rocky Mountains
 - Poor: 90th percentile = 75% fines Southern Rocky Mountains



Talk Outline

- Compare and contrast common benchmark methods for conductivity and fine sediment
- How do the concepts of accuracy and precision apply to benchmark methods?
- How do we rank stressors given differences in benchmark method accuracy and precision across indicators and agencies?

Accuracy and Precision

General

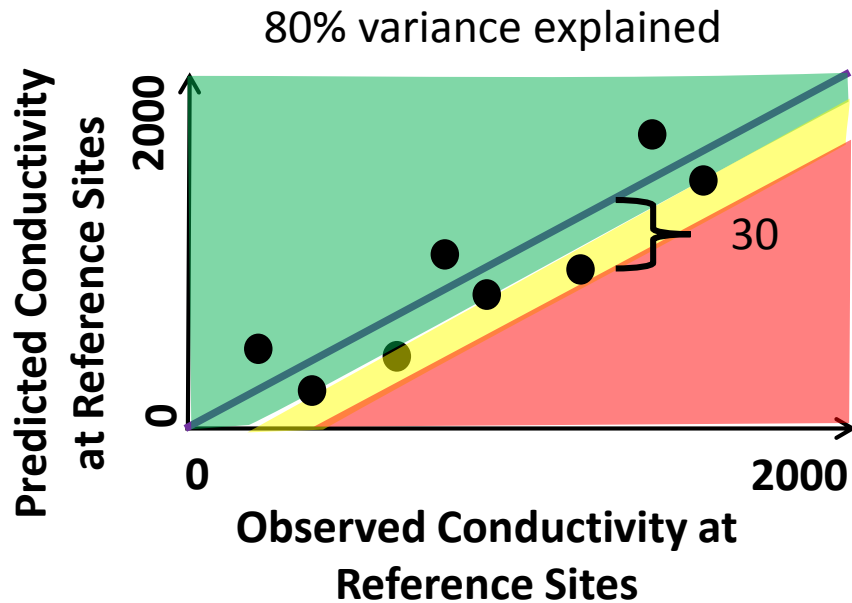
- Accuracy – how close a measurement or prediction is to the true value
- Precision – how repeatable values are

Benchmark Method

- Accuracy – how well the method “predicts” what reference condition should be for a given site
- Precision – how much variation among reference sites remains after attempting to account for natural variation

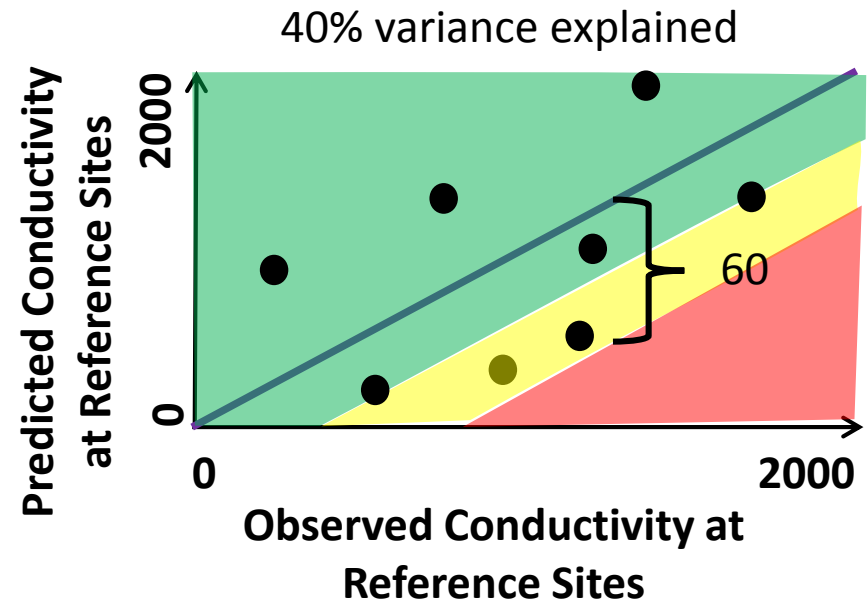
Accuracy and Precision: Empirical Models

Model quality matters. Imprecise models are less protective.



Poor: Predicted value + 95th percentile of residuals

$$\text{Poor: } 500 + 30 = 530 \mu\text{S/cm}$$



$$\text{Poor: } 500 + 60 = 560 \mu\text{S/cm}$$

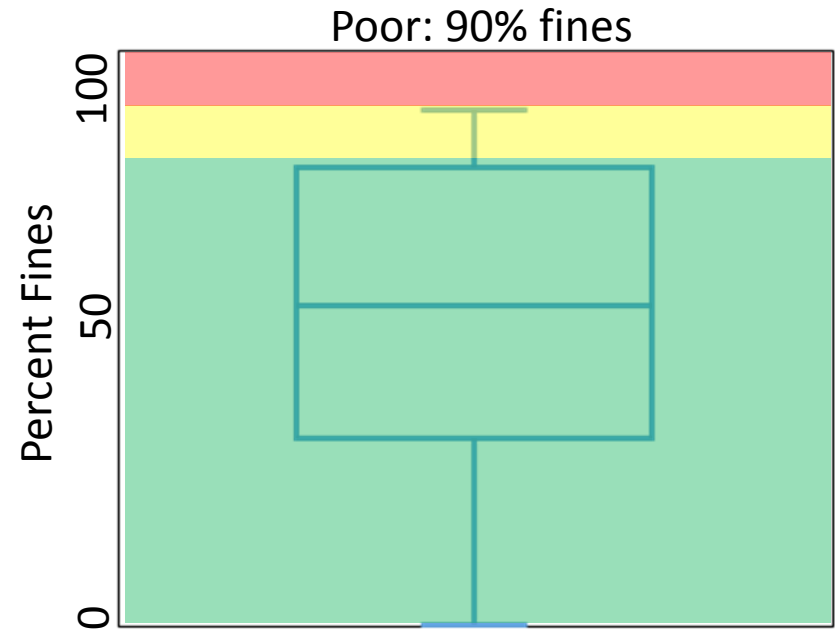
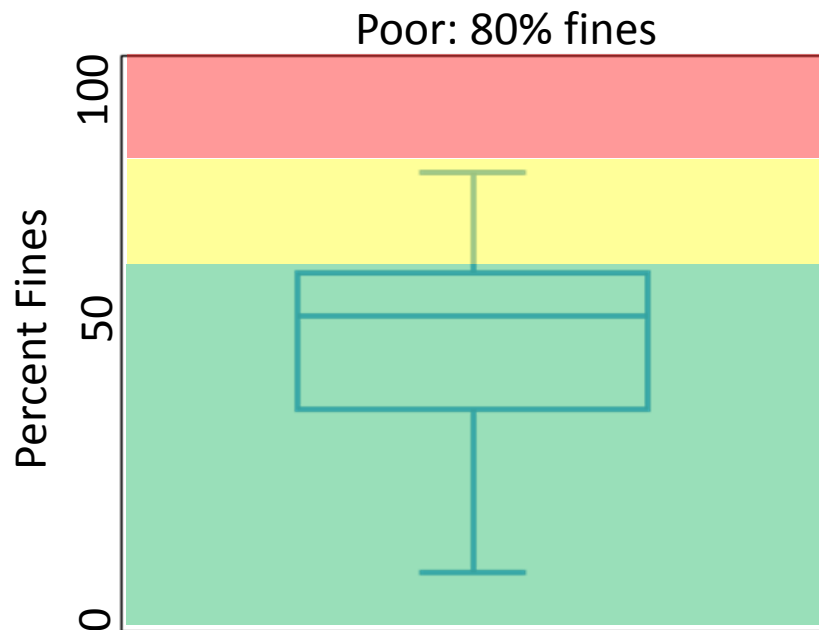
Accuracy and Precision: Percentiles of Reference

Accuracy

- Distribution may accurately represent population but one benchmark is used for all sites; benchmark will be accurate for some sites but not others

Precision

- Larger the spread of the distribution the less precise and the less protective



Can't compare accuracy and precision across methods

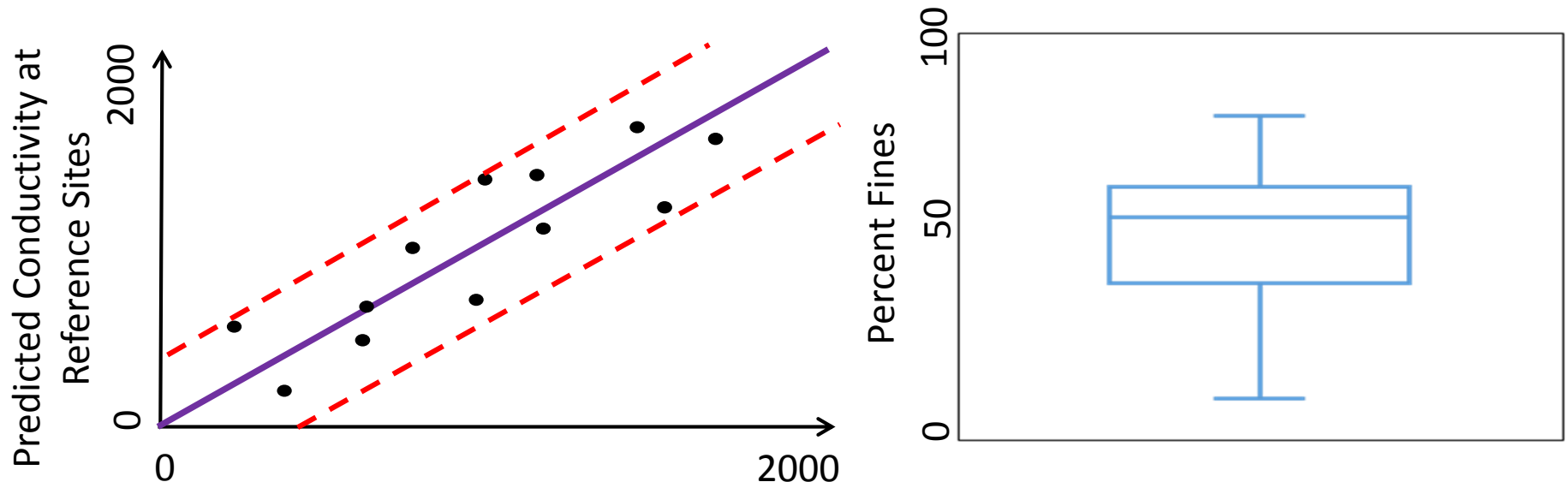
Models

- Have known accuracy and precision
 - R^2 and RMSE

Percentiles of Regional Reference Site Values

- No equivalent metrics to compare to models

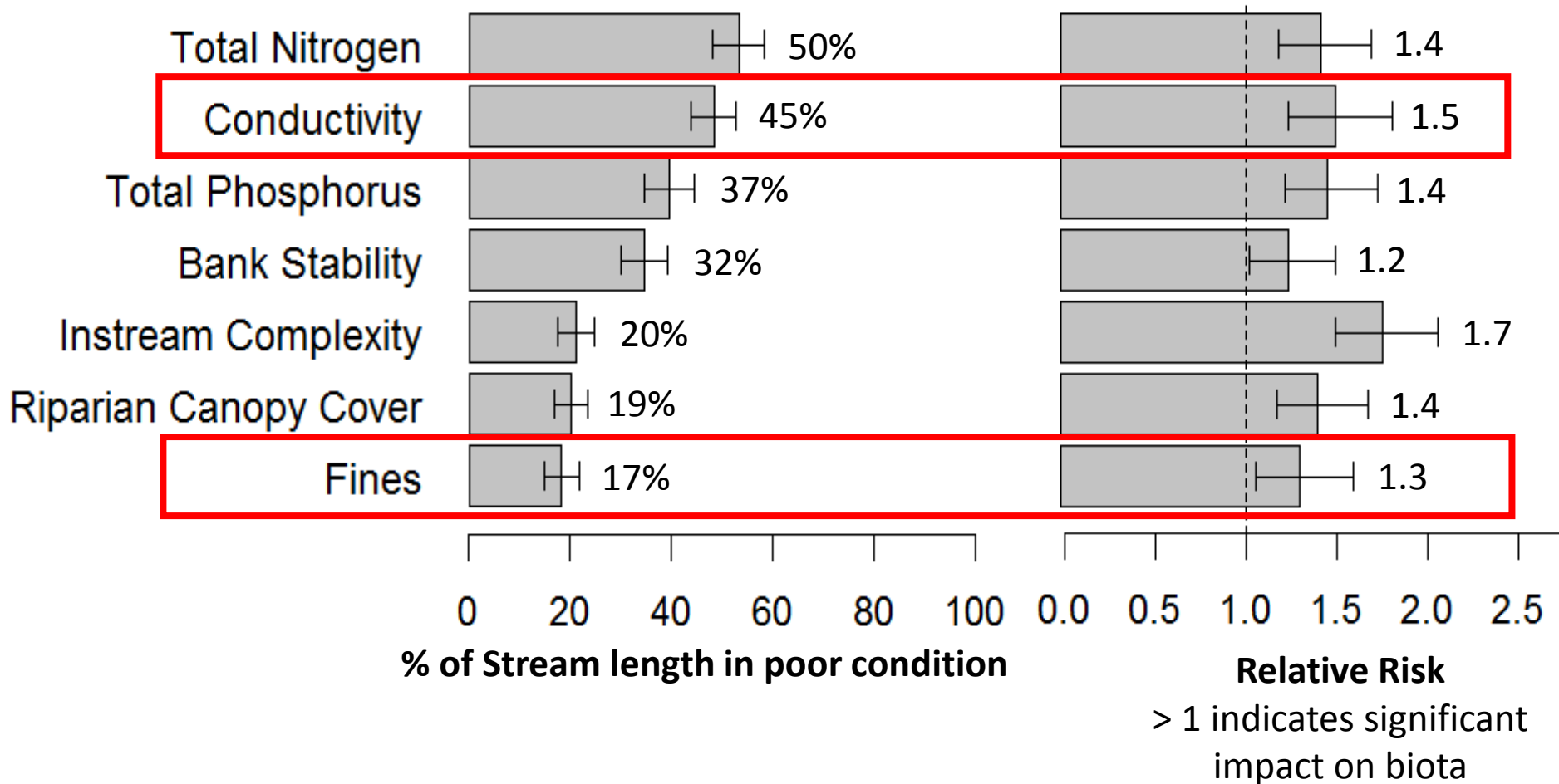
Variance of model residuals is different than variance of reference distribution



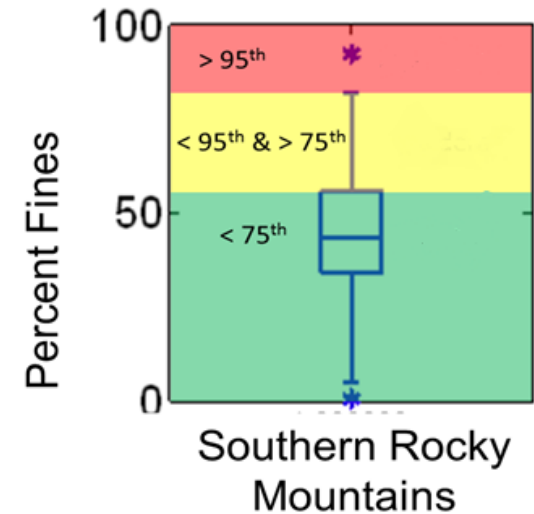
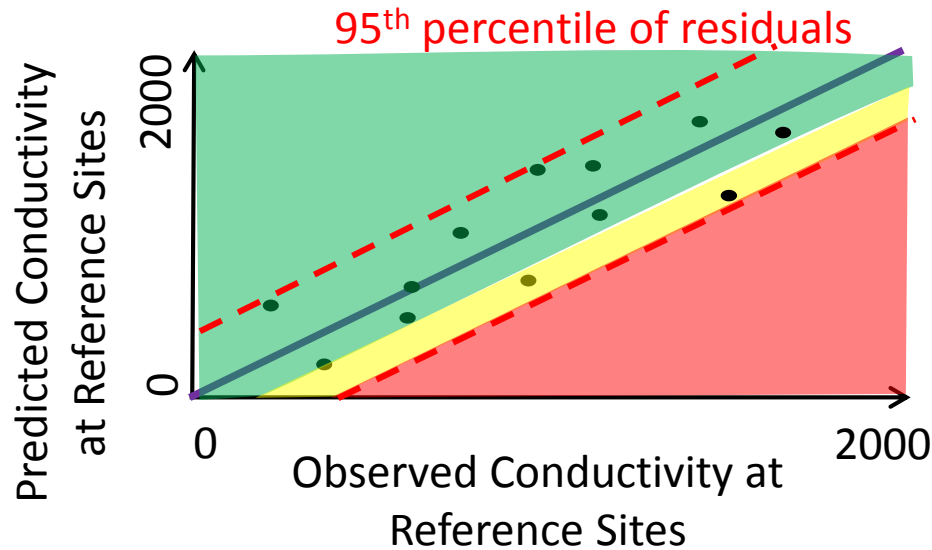
Talk Outline

- Compare and contrast common benchmark methods for conductivity and fine sediment
- How do the concepts of accuracy and precision apply to benchmark methods?
- How do we rank stressors given differences in benchmark method accuracy and precision across indicators and agencies?

Consider biological relevance



Choose percentiles that balance over vs. under-protection



- Over-protection – flagging more sites in poor condition than there actually are
- Under-protection – not flagging sites that actually are in poor condition
- Don't just automatically use 95th and 75th percentiles to select benchmarks
- The lower the percentile the stricter and more protective the benchmark

Use modeling approaches where possible

Indicator	Method	Agency
Conductivity	Model (Olson and Hawkins 2012)	BLM
	Best professional judgement	EPA and most states
% Fines	Percentiles of regional reference	BLM
	RBS Model (Kaufmann et al. 2018)	EPA
	Biological response	States

Barriers

- Model availability and scale
- Model precision and accuracy
- Communication challenges

Summary

- Benchmark methods differ in how well they account for natural environmental gradients
- All methods are biased but some are more susceptible to over or under protection
- Different benchmark methodologies lead to drastically different results and confound comparison among indicators and agencies
- Some best practices to consider
 - Biological relevance
 - Balancing over vs. under-protection
 - Available modeling options

Needs

Technical

- More site-specific models of physical habitat
- More interagency networks of reference sites

Communication

- Standard terminology of
 - Benchmark methods
 - Method accuracy and precision concepts
- Non-technical explanations of models
- Guidance for managers to adjust benchmarks depending on the method and application